

Ergodicity and diffusivity of Markovian order book models: a general framework

Weibing Huang^{1,2} and Mathieu Rosenbaum¹

¹ LPMA, University Pierre et Marie Curie (Paris 6)

² Kepler-Cheuvreux

May 20, 2015

Abstract

We present a general Markovian framework for order book modeling. Through our approach, we aim at providing a tool enabling to get a better understanding of the price formation process and of the link between microscopic and macroscopic features of financial assets. To do so, we propose a new method of order book representation, and decompose the problem of order book modeling into two sub-problems: dynamics of a continuous-time double auction system with a fixed reference price; interactions between the double auction system and the reference price movements. State dependency is included in our framework by allowing the order flow intensities to depend on the order book state. Furthermore, contrary to most existing models, the impact of the order book updates on the reference price dynamics is not assumed to be instantaneous. We first prove that under general assumptions, our system is ergodic. Then we deduce the convergence towards a Brownian motion of the rescaled price process.

1 Introduction

Nowadays, most financial exchanges use a limit order book (LOB) mechanism. In these order-driven markets, market participants send their buy and sell orders to a continuous-time double auction system, where orders are matched according to their price and time priority. Understanding the LOB dynamics is one of the fundamental issues in the field of market microstructure and leads to many interesting applications in optimal execution, design of electronic trading algorithms, minimization of market impact costs, short-term predictions and regulation. In the recent years, many works have been devoted to the description of order book dynamics. Order book models can be essentially divided into two types: economic models, where one focuses on the behaviors of individual agents and their optimal decisions, see for example Parlour (1998), Foucault (1999) and Roşu (2009); statistical models, where the order flows are seen as random processes, see Smith, Farmer, Gillemot, and Krishnamurthy (2003), Cont, Stoikov, and Talreja (2010), Abergel and Jedidi (2011), Cont and De Larrard (2013), Lakner, Reed, and Stoikov (2013), Lachapelle, Lasry, Lehalle, and Lions (2014), Bayer, Horst, and Qiu (2015) and Abergel and Jedidi (2015). With the notable exception of Abergel and Jedidi (2015), where the authors consider the case of Hawkes-type dynamics, these models usually assume Poisson flows

for the order arrival processes. Such assumption is mainly made for technical reasons, since it is well-known that it is not consistent with market data. In Huang, Lehalle, and Rosenbaum (2015), the authors propose to replace the Poisson assumption by a state-dependent approach where the intensities of the flows depend on the state of the LOB. This model, called “Queue-reactive” model, provides new insights for the order book dynamics, such as market participants behaviors conditional on different states of the order book, the LOB’s asymptotic form and the bid-ask spread distribution. It is also a very relevant tool for practitioners in the perspective of transaction cost analysis of complex trading algorithms. In this paper, we aim at extending the Queue-reactive model to a more general framework, in which most of the existing statistical models can be included (up to minor modifications). Our goal is to give some theoretical results on the system’s ergodicity as well as the asymptotic scaling limit of the price process.

In the LOB, price levels are discretized by a minimum price change unit called the tick value (denoted by α). Market participants can place their buy/sell orders at any level which is a multiple of the tick value and these orders will either stay in the LOB (a buy order with price lower than the current best ask price, or a sell order with price higher than the current best bid price, this type of orders being called “limit order”), or be matched with the existing orders in the LOB (this type of orders being called “market order”). The LOB, as its name suggests, is composed of all unmatched limit orders and can be seen as a (rough) approximation of the current microstructural supply and demand on the different price levels.

Current statistical models differ in their way of representing the LOB. In Cont, Stoikov, and Talreja (2010), the price grid is supposed to be finite ($n_{min}\alpha, \dots, n_{max}\alpha$), and the LOB is represented by a $n_{max} - n_{min} + 1$ dimensional vector that records the buying/selling quantities at each of these price levels. In such representation, the different limits are indexed by their absolute price level. In practice, to cover the intra-day price range $[p_{min}, p_{max}]$, the dimension of the state space have to be at least $\frac{p_{max} - p_{min}}{\alpha} + 1$ ¹, which is typically a very large number. Another way of representing the order book state is to use the relative indexing method. Following ideas from the Zero-intelligence model of Smith, Farmer, Gillemot, and Krishnamurthy (2003), Abergel and Jedidi (2011) propose to use the best bid and the best ask prices as two reference prices to index the limits. In that case, the LOB is made of the following elements: the two reference prices $p_{bestbid}$ and $p_{bestask}$, and the limits around them which are two K dimensional vectors a (for the ask side) and b (for the bid side). The vector $a = [a_1, \dots, a_K]$ records the limit sizes at the price levels $[p_{bestbid} + \alpha, \dots, p_{bestbid} + K\alpha]$, while the vector $b = [b_1, \dots, b_K]$ records the limit sizes at the price levels $[p_{bestask} - \alpha, \dots, p_{bestask} - K\alpha]$ ². In practice, observing a market depth of five ticks is usually considered enough for most trading purposes. Consequently, for a typical stock with spread size of order five ticks, the value of K should be generally of order 10 so that essential information from the LOB is captured. Thus, the use of these two reference prices reduces significantly the dimensionality of the state space. Note that in this representation, the index of a limit at a given price level is no longer constant. Therefore, appropriate boundary conditions must be defined to deal with price changes.

In this paper, we propose an original representation of the order book, using only one reference

¹Of course if we choose $n_{min} = 1$ and $n_{max} = \infty$, we have a complete description of all the available buying/selling offers in the whole price grid $\alpha\mathbb{N}$.

²Note that in this representation, there can be overlaps between the vector a and b when the bid-ask spread is larger than one tick.

price which is not necessarily directly observable from the order book state. We view this reference price as sort of market consensus about the underlying “efficient” price used by market participants when making their trading decisions. We keep K limits on each side of the reference price and the LOB is fully described by a $2K + 1$ dimensional vector, which is then modeled by a continuous-time Markov jump process. The use of this unique reference price gives us a lot of flexibility when modeling the order book. Since the reference price is no longer directly determined by the order book state, we can differentiate two types of jumps in the Markov process: pure order book state jumps (jumps for which the reference price stays constant) and common jumps (jumps for which a reference price change is involved). For large tick assets³, such decomposition is proved to be very relevant when studying the conditional dependences between the dynamics of the LOB and its state, see Huang, Lehalle, and Rosenbaum (2015). Moreover, in this framework, we are able to easily incorporate exogenous price movements into the order book dynamics. This can be simply done adding a reference price jump component which is independent of the order book state.

At the high frequency scale, the LOB state is one of the two public information that are accessible to traders and their automates (the other being the history of the order flows). Thus it plays a very important role in their trading decisions. In our framework, the LOB is assumed to be a continuous-time Markov jump process, and the influence of the LOB state on the incoming flows is modeled through a state-dependent infinitesimal generator matrix for the jump process. Indeed, in practice, traders essentially rely on information deduced from the current LOB state when deciding to send an order at a specific price level. Various simplifying assumptions on the information set used by traders can be considered in our framework in order to facilitate the empirical studies. The index of a limit, for example, is probably one of the most important elements in their decision process, as it gives the distance between the target price and the reference price. Influence of other variables, such as the target limit’s size, its relative distance to the current best offer queues and the size of its opposite queue is studied in Huang, Lehalle, and Rosenbaum (2015), and are shown to also have non-negligible effects on the dynamics of the order flows.

Under appropriate assumptions, the Queue-reactive model can be easily estimated using empirical data. It provides many interesting new insights on the origin of some micro-structural properties, such as the stylized empirical distribution of the LOB state. It has been shown in Bouchaud, Mézard, and Potters (2002) that there exist some regularities in the order book’s empirical form, that is the average value of the LOB state (a $2K$ dimensional vector in our model). From a theoretical point of view, these regularities are closely linked with the notion of ergodicity of the LOB system (the exact definition of ergodicity will be given in Section 3): Ergodicity ensures the convergence of the LOB state distribution towards an invariant probability measure. Thus the stylized form observed on market data might be explained by a law of large number type phenomenon for this invariant distribution. This hypothesis is supported by empirical studies in Huang, Lehalle, and Rosenbaum (2015), in which the authors compare the theoretical asymptotic distributions in our model with empirical estimations, and show that they are very close. In Huang, Lehalle, and Rosenbaum (2015), some assumptions are made to ensure the LOB system’s ergodicity. In this paper, we want to generalize them in an extended framework where the volume of the orders is no longer constant and the influence of the order book state on the dynamics of the reference price may not be instantaneous.

³A large tick asset is defined as an asset whose bid-ask spread is almost always equal to one tick.

Another important element in order book modeling is the asymptotic behavior of the price. Such analysis is very relevant as it provides useful insights on the price formation process, and links the dynamics at the microscopic level with macroscopic features of the asset, such as its volatility. We prove that in our framework, the rescaled reference price process converges to a Brownian motion. An expression for the macroscopic volatility in terms of the flow rates is derived using a functional central limit theorem together with the strong mixing property of the price increments, in the spirit of Abergel and Jedidi (2011).

The paper is organized as follows. In Section 2, we set up the general Markovian framework. Section 3 discusses the ergodicity properties of the model. The diffusive limit of the rescaled price process is stated in Section 4. Finally, in Section 5, we give some specific models which can be seen as particular cases of our framework. The technical proofs are gathered in an appendix.

2 A general Markovian framework

2.1 Representation of the order book

In our framework, the order book is made of two elements: the center position: a certain reference price (note that here the center position is not necessarily the mid price) and the shape of the book (the queue sizes around the reference price), see Figure 1 for an example. The center position, denoted by $p_{ref} \in \{n\alpha + 0.5\alpha, n \in \mathbb{Z}\}^4$, can be seen as the current consensus price level and is used to index the limits. We write $Q_{\pm i}$ for the queue at the price level $p_{ref} \pm (i - 0.5)\alpha$ and denote its size by $q_{\pm i}$ and its price by p_i . Then the total order book's shape at time t is an infinite vector $q(t) = [\dots, q_{-k}(t), \dots, q_{-1}(t), q_1(t), \dots, q_k(t), \dots]$, with $q_i \in \mathbb{Z}$, and $|q_i|$ is the number of orders at the limit Q_i ($q_i < 0$ if these orders are bid orders and $q_i > 0$ if they are ask orders. Note that in such a representation, one can have $q_i \geq 0$ or $q_i < 0$ for all i). The LOB information at time t is therefore fully represented by $X(t) := (q(t), p_{ref}(t))$, $t \geq 0$. To restrict the value of $X(t)$, we consider only K limits on each side, and thus have now $q(t) = [q_{-K}(t), \dots, q_{-1}(t), q_1(t), \dots, q_K(t)]$ and $X(t) \in \mathbb{Z}^{2K} \times \alpha(0.5 + \mathbb{Z})$.

The state space of $q(t)$ is actually smaller than \mathbb{Z}^{2K} . More specifically, let us define

$$\begin{aligned} i_{bestbid}(q) &= \max(-K - 1, \sup\{i | q_i < 0\}) \\ i_{bestask}(q) &= \min(K + 1, \inf\{i | q_i > 0\}). \end{aligned}$$

The state space Ω of $q(t)$ is defined as the set of all $q \in \mathbb{Z}^{2K}$, such that: for all $i \in \{-K, \dots, K\}$, $q_i \leq 0$ if $i \leq i_{bestbid}(q)$ and $q_i \geq 0$ if $i \geq i_{bestask}(q)$.

2.2 Dynamics of the order book

In our general framework, we model the LOB vector $X(t)$ as a continuous-time Markov jump process, whose infinitesimal generator matrix \mathcal{Q} will be given in Equation (4). We differentiate two types of jumps in the order book dynamics: pure order book state jumps (for which p_{ref} stays constant) and common jumps (for which a change in the value of p_{ref} is involved).

⁴For the generality of the framework, we allow for negative prices, however, in practice, the model should of course be used over a reasonably small time interval so that prices remain positive.



Figure 1: Limit order book

Pure order book state jump

There are three⁵ types of orders that interact directly with the order book and trigger pure order book state jumps:

- Limit orders: insertion of a new order in the order book (a buy order at a lower price than the best ask price, or a sell order at a higher price than the best bid price).
- Cancellation orders: cancellation of an already existing order in the order book.
- Market orders: consumption of available liquidity (a buy or sell order at the best available price).

In the seminal work of Smith, Farmer, Gillemot, and Krishnamurthy (2003), the arrival times of the above three types of orders at different price levels are assumed to be mutually independent and exponentially distributed. Furthermore, each order has unit size. In our approach, the size of the jumps, which represents the amount of volume inserted to/removed from the LOB for a given event, is random. Moreover, the arrival rate of a given jump is assumed to be function of the index of the target price, the current LOB state vector $q(t)$, the direction of the jump and its size. That is, for any $q, q' \in \Omega$ ($q \neq q'$), $p \in \alpha(0.5 + \mathbb{Z})$, $n \in \mathbb{N}^+$ and any $e_i = (a_{-K}, \dots, a_i, \dots, a_K)$ ($a_j = 0$ for $j \neq i$ and $a_i = 1$), we have in cases where $q + ne_i \in \Omega$ and $q - ne_i \in \Omega$:

$$\begin{aligned}
 \mathcal{Q}_{(q,p),(q+ne_i,p)} &= f_i(q, n) \\
 \mathcal{Q}_{(q,p),(q-ne_i,p)} &= g_i(q, n) \\
 \tilde{\mathcal{Q}}_{(q,p),(q',p)} &= 0, \text{ otherwise,}
 \end{aligned} \tag{1}$$

where the f_i and g_i are $2K$ functions: $\Omega \times \mathbb{N}^+ \rightarrow \mathbb{R}^+$.

⁵Four if we also consider the modification orders. We view modification orders as a combination of a cancellation and an insertion order that arrive in a very short time interval.

Note that in (1), $f_i(q, n)$ and $g_i(q, n)$ have different meanings for different i and q . For example, when $i \geq i_{bestask}$, $f_i(q, n)$ represents the arrival rate of sell limit orders of size n , and $g_i(q, n)$ the sum of the rate of cancellations of size n and the arrival rate of market buy orders of size n . When $i \leq i_{bestbid}$, the role of $f_i(q, n)$ and $g_i(q, n)$ are switched. Note also that $q \pm ne_i$ is not always in the state space Ω even when $q \in \Omega$. Thus some values of the functions f_i and g_i are not needed in Equation (1) and assumed to be equal to zero. Furthermore, so that there is no absorbing state, we assume

$$\sum_i \sum_n (f_i(q, n) + g_i(q, n)) > 0.$$

Common jumps

The reference price p_{ref} can be viewed as a consensus value on the “efficient” price and takes discretized values in $\alpha(0.5 + \mathbb{Z})$. In practice, this reference price is built based on two sets of information: the current state of the LOB and the historical order flows. We have in mind that p_{ref} moves in a Markovian manner, so its dynamics depends on the present information only, that is the current state of the LOB. In our framework, we restrict the price jump size to one tick at each time. We use two functions $u, d : \Omega \rightarrow \mathbb{R}^+$ to describe respectively the rate of positive and negative jumps:

$$\begin{aligned} \sum_{q' \in \Omega} \mathcal{Q}_{(q,p),(q',p+\alpha)} &= u(q) \\ \sum_{q' \in \Omega} \mathcal{Q}_{(q,p),(q',p-\alpha)} &= d(q) \\ \sum_{q' \in \Omega} \mathcal{Q}_{(q,p),(q',p \pm n\alpha)} &= 0, \text{ for } n \geq 2. \end{aligned} \tag{2}$$

To understand Equation (2), let us first consider the following simple example where the LOB state information is summarized by the difference between the current value of p_{ref} and the mid price p_{mid} :

Example 1.

$$\begin{aligned} i_{mid} &= (i_{bestbid} + i_{bestask})/2, \\ u(q) &= \theta_0 + \theta_1 \max(0, \alpha(i_{mid} - 0.5)) \\ d(q) &= \theta_0 + \theta_1 \max(0, -\alpha(i_{mid} + 0.5)), \end{aligned}$$

with $\theta_0 \geq 0$ and θ_1 a positive constant representing the intensity of the adjustment of p_{ref} towards p_{mid} .

In the above example, we assume that the reference price jump rate depends on the deviation of the current value of p_{ref} from p_{mid} . Indeed, p_{mid} is often considered an approximation of the LOB center implied by its current state. We may also include other LOB information such as the available quantities at $Q_{i_{bestbid}}$ and $Q_{i_{bestask}}$ when defining $u(q)$ and $d(q)$. Such additional variables increase the complexity of our model but make it more realistic.

Remark 2.1. In the already mentioned Queue-reactive model, changes of p_{ref} are triggered by the order book events that modify p_{mid} , while in Equation (2), they are driven by the order book state. Although the Queue-reactive model cannot exactly be seen as a particular case of the framework presented in this paper (see Section 5), most of the theoretical results shown in Sections 3 and 4 still hold (with some minor modifications in some assumptions). In particular, one can prove the diffusive limit of the reference price in the Queue-reactive model applying almost the same method as that used here.

As soon as p_{ref} changes, the value of q_i switches immediately to the value of one of its neighbors (right if p_{ref} increases, left if it decreases). As we keep only K limits on each side, two boundary distributions π_{-K} and π_K are introduced for generating the new queue sizes at Q_{-K} (when p_{ref} decreases) and Q_K (when p_{ref} increases). To possibly incorporate external information, we moreover assume that with probability θ^{reinit} , the LOB state vector $q(t)$ is redrawn from some distribution (π^{inc} if p_{ref} increases, π^{dec} if p_{ref} decreases) when p_{ref} changes. As shown in Huang, Lehalle, and Rosenbaum (2015), models where price dynamics are purely endogenous, driven by order flows only, are usually not able to reproduce some of the important macroscopic features of prices, such as the volatility. Thus the parameter θ^{reinit} can be understood as the percentage of price changes due to exogenous information, in which case market participants readjust very quickly their order flows around the new reference price, as if a new state of the LOB was drawn from its (invariant) distribution (ergodicity conditions are discussed in the next section).

For $q \in \Omega$, we write $q^+ = [q_{-K+1}, \dots, q_{-1}, q_1, \dots, q_K]$, $q^- = [q_{-K}, \dots, q_{-1}, q_1, \dots, q_{K-1}]$, $[q^+, l] = [q_{-K+1}, \dots, q_{-1}, q_1, \dots, q_K, l]$ and $[l, q^-] = [l, q_{-K}, \dots, q_{-1}, q_1, \dots, q_{K-1}]$. Under the above assumptions, we have for $l \in \mathbb{Z}$ and $q, q', q'' \in \Omega$ such that $q'^+ \neq q^+$ and $q''^- \neq q^-$:

$$\begin{aligned} \mathcal{Q}_{(q,p),([q^+,l],p+\alpha)} &= (1 - \theta^{reinit})u(q)\pi_K(l) + \theta^{reinit}u(q)\pi^{inc}([q^+,l]) \\ \mathcal{Q}_{(q,p),(q',p+\alpha)} &= \theta^{reinit}u(q)\pi^{inc}(q') \\ \mathcal{Q}_{(q,p),([l,q^-],p-\alpha)} &= (1 - \theta^{reinit})d(q)\pi_{-K}(l) + \theta^{reinit}d(q)\pi^{dec}([l,q^-]) \\ \mathcal{Q}_{(q,p),(q'',p-\alpha)} &= \theta^{reinit}d(q)\pi^{dec}(q''). \end{aligned} \tag{3}$$

The infinitesimal generator matrix of $X(t)$

Equations (1), (2) and (3) give a complete description of the infinitesimal generator matrix \mathcal{Q} of the process $X(t)$, which is summarized in the following assumption.

Assumption 2.1. Let $q, q', q'', \tilde{q} \in \Omega$, $p, \tilde{p} \in \alpha(0.5 + \mathbb{Z})$, $n \in \mathbb{N}^+$, $l \in \mathbb{Z}$ be such that $q'^+ \neq q^+$ and $q''^- \neq q^-$. The process $X(t)$ is an irreducible Markov jump process with aperiodic embedded chain whose infinitesimal generator matrix \mathcal{Q} is of the following form (with $2K$ functions f_i, g_i :

$\Omega \times \mathbb{N}^+ \rightarrow \mathbb{R}^+$ and two functions $u, d : \Omega \rightarrow \mathbb{R}^+$:

$$\begin{aligned}
\mathcal{Q}_{(q,p),(q+ne_i,p)} &= f_i(q, n) \\
\mathcal{Q}_{(q,p),(q-ne_i,p)} &= g_i(q, n) \\
\mathcal{Q}_{(q,p),([q^+, l], p+\alpha)} &= (1 - \theta^{reinit})u(q)\pi_K(l) + \theta^{reinit}u(q)\pi^{inc}([q^+, l]) \\
\mathcal{Q}_{(q,p),(q', p+\alpha)} &= \theta^{reinit}u(q)\pi^{inc}(q') \\
\mathcal{Q}_{(q,p),([l, q^-], p-\alpha)} &= (1 - \theta^{reinit})d(q)\pi_{-K}(l) + \theta^{reinit}d(q)\pi^{dec}([l, q^-]) \\
\mathcal{Q}_{(q,p),(q'', p-\alpha)} &= \theta^{reinit}d(q)\pi^{dec}(q'') \\
\mathcal{Q}_{(q,p),(q,p)} &= - \sum_{(\tilde{q}, \tilde{p}) \in \Omega \times \alpha(0.5 + \mathbb{Z}), (\tilde{q}, \tilde{p}) \neq (q,p)} \mathcal{Q}_{(q,p),(\tilde{q}, \tilde{p})}, \\
\mathcal{Q}_{(q,p),(\tilde{q}, \tilde{p})} &= 0, \text{ otherwise.}
\end{aligned} \tag{4}$$

Note that under Assumption (2.1), the dynamics of the process $X(t)$ is invariant under translations of the LOB center position: its infinitesimal matrix generator \mathcal{Q} satisfies:

$$\mathcal{Q}_{(q^1, p^1), (q^2, p^1 + \beta)} = \mathcal{Q}_{(q^1, p^2), (q^2, p^2 + \beta)},$$

for any $q^1, q^2 \in \Omega$, $p^1, p^2 \in \alpha(0.5 + \mathbb{Z})$ and $\beta \in \alpha\mathbb{Z}$. One can also remark that in our framework, the order book state process $q(t)$ alone is also a continuous-time Markov jump process, whose ergodicity is discussed in the next section.

2.3 Comparison with existing models

The first major difference between our approach and the existing Markovian models in the literature is the introduction of state dependency in the order book dynamics. Most of the current order book models follow the “Zero-intelligence” framework, using Poisson flows for the processes of order arrivals. The Poisson assumption is clearly unrealistic, see for example the empirical results in Huang, Lehalle, and Rosenbaum (2015). In our framework, we propose to incorporate the strategic behaviors of market participants via a mean-field game approach, assuming their decisions depend on an underlying “efficient” price p_{ref} and on the LOB state vector $q(t)$. Note also that Equation (1) allows us to have jumps of random size in the order book’s shape, while a constant jump size is often assumed in the other existing models.

We also introduce a new method of LOB representation using one unique reference price p_{ref} . Most models use the best bid and best ask prices as two reference prices for indexing the buy and sell limits. In such models, these reference prices are directly determined by the order book state. In particular, changes in the order book state are immediately carried on the values of these prices. In our framework, p_{ref} is not necessarily deduced from the order book state. Therefore, we can assume that changes in the order book state affect the value of p_{ref} with some delay rate (the functions u and d introduced in Equation (2)). Thanks to this original representation, we can naturally decompose the order book dynamics into two parts: a continuous-time multidimensional queuing system (Equation (1)) and the dynamics of its center, that is the reference price (Equations (2) and (3)). Compared with the Queue-reactive model introduced in Huang, Lehalle, and Rosenbaum (2015), p_{ref} is no longer constrained within the bid-ask spread. This desirable feature gives us the possibility of separating the

exogenous and endogenous parts in the price dynamics by choosing appropriate price jump rate functions u and d . For example, with the functions u and d of Example 1, one can interpret θ_0 as the exogenous part in the dynamics of p_{ref} and θ_1 as the intensity of the endogenous effects driving p_{ref} towards the current mid price level.

3 Ergodicity

In this section, we discuss ergodicity properties in our framework. To do so, we make some additional assumptions on the functions f_i , g_i , u and d .

Let

$$P_t(x, A) := \mathbb{P}[Y(t) \in A | Y(0) = x]$$

be the transition probability at time t of a continuous or discrete-time Markov process Y with state space \mathcal{Y} . In this work, we say that the process Y is *V-uniformly ergodic* if there exists a coercive⁶ function $V > 1$, an invariant distribution π , $r \in (0, 1)$ and $R > \infty$ such that for any $x \in \mathcal{Y}$ and $t \in \mathbb{R}^+$ (or \mathbb{N}^+ in discrete-time),

$$\|P_t(x, \cdot) - \pi(\cdot)\|_V \leq Rr^t V(x), \quad (5)$$

where we write $\|\cdot\|_V$ for the V -norm of a signed measure, see Meyn and Tweedie (1993, 2009). In continuous time, the main idea to prove such property is to design an appropriate Lyapunov function $V: \mathcal{Y} \rightarrow (1, \infty)$, on which the following negative drift condition is satisfied for some $\gamma > 0$ and $B > 0$:

$$\begin{aligned} \mathcal{Q}V(y) &:= \sum_{y' \neq y} \mathcal{Q}_{yy'}[V(y') - V(y)] \\ &\leq -\gamma V(y) + B. \end{aligned}$$

Then by Theorem 6.1 in Meyn and Tweedie (1993), the Markov process Y is non-explosive and V -uniformly ergodic. Furthermore, by Theorem 4.2 in Meyn and Tweedie (1993) it is positive Harris recurrent. Note that the same kind of method is used in Abergel and Jedidi (2011) in order to show ergodicity properties of Zero-intelligence models.

As mentioned in the introduction, the LOB's ergodicity implies here the existence of a unique invariant distribution for the state vector q . This is relevant for explaining the stylized empirical distribution of the LOB state. Mostly, as we will see in Section 4, the ergodicity analysis is the basis for proving the diffusive limit of the reference price process.

3.1 When p_{ref} stays constant

We first discuss the V -uniform ergodicity of the process $q(t)$ when assuming $u(q) = d(q) = 0$ in Equation (4). Recall that the unused values of $f_i(q, n)$ and $g_i(q, n)$ in the definition of the queue dynamics, that is when $q \pm ne_i \notin \Omega$, are set to zero. With the convention $0/0 = 0$, we

⁶ $|V(x)| \rightarrow +\infty$ as $\|x\| \rightarrow \infty$.

define

$$\begin{aligned}
f_i^*(q) &:= \sum_n f_i(q, n) \\
g_i^*(q) &:= \sum_n g_i(q, n) \\
l_i(q, n) &:= \frac{f_i(q, n)}{f_i^*(q)} \\
k_i(q, n) &:= \frac{g_i(q, n)}{g_i^*(q)} \\
G^{f,i,q}(z) &:= \sum_{n=1}^{\infty} z^n l_i(q, n) \\
G^{g,i,q}(z) &:= \sum_{n=1}^{\infty} z^n k_i(q, n).
\end{aligned}$$

Thus, when $f_i^*(q) > 0$ (resp. $g_i^*(q) > 0$), $l_i(q, \cdot)$ (resp. $k_i(q, \cdot)$) is a probability measure on \mathbb{N}^+ with moment-generating function $G^{f,i,q}(z)$ (resp. $G^{g,i,q}(z)$). We make the four following assumptions.

Assumption 3.1. *For any order book state q and any $i \geq i_{bestask}$, $g_i(q, n) = 0$ for any $n > q_i$ and for any order book state q and any $i \leq i_{bestbid}$, $f_i(q, n) = 0$ for any $n > -q_i$.*

Assumption 3.2. *There exists $z^* > 1$ such that for any q and i , $G^{f,i,q}(z^*) < \infty$ and $G^{g,i,q}(z^*) < \infty$. Furthermore, there exists $L > 0$ such that for any i ,*

$$\overline{\lim}_{z \rightarrow 1^+} \sup_q [f_i^*(q) G^{f,i,q}(z) \mathbf{1}_{i > i_{bestbid}} + g_i^*(q) G^{g,i,q}(z) \mathbf{1}_{i < i_{bestask}}] < L.$$

Assumption 3.3. *There exist $r > 0$ and $U > 1$ such that*

$$\begin{aligned}
\overline{\lim}_{z \rightarrow 1^+} \sup_{(q,i): q_i > U, i \geq i_{bestask}} [f_i^*(q) - g_i^*(q) \frac{1 - G^{g,i,q}(z^{-1})}{G^{f,i,q}(z) - 1}] &< -r \\
\overline{\lim}_{z \rightarrow 1^+} \sup_{(q,i): q_i < -U, i \leq i_{bestbid}} [g_i^*(q) - f_i^*(q) \frac{1 - G^{f,i,q}(z^{-1})}{G^{g,i,q}(z) - 1}] &< -r.
\end{aligned} \tag{6}$$

Assumption 3.4. *For any $z > 1$,*

$$\begin{aligned}
B_f(z) &:= \inf_{(q,i): q_i > U, i \geq i_{bestask}} (G^{f,i,q}(z) - 1) > 0 \\
B_g(z) &:= \inf_{(q,i): q_i < -U, i \leq i_{bestbid}} (G^{g,i,q}(z) - 1) > 0.
\end{aligned}$$

To understand the practical meaning of these assumptions, let us consider the following example where the pure order book state jumps are assumed to have constant size equal to one. In such situation, the four assumptions above can be rewritten as follows and are much easier to interpret.

Example 2. *LOB model with constant order size.*

- For $n \geq 2$, $f_i(q, n) = g_i(q, n) = 0$ for any $q \in \Omega$.
- There exists $L > 0$ such that for any $i \in \{-K, \dots, K\}$ and $q \in \Omega$,

$$f_i(q, 1)\mathbf{1}_{i > i_{bestbid}} + g_i(q, 1)\mathbf{1}_{i < i_{bestask}} < L.$$

- There exist $r > 0$ and $U > 1$ such that

$$\begin{aligned} \sup_{(q,i): q_i > U, i \geq i_{bestask}} [f_i(q, 1) - g_i(q, 1)] &< -r \\ \sup_{(q,i): q_i < -U, i \leq i_{bestbid}} [g_i(q, 1) - f_i(q, 1)] &< -r. \end{aligned}$$

Basically, Assumption 3.1 says that a bid/ask limit cannot become an ask/bid limit in a single queue update event, that is the queue size cannot revert its sign in a single jump⁷. From Example 2, we see that Assumption 3.2 essentially states that the total intensity of the order insertion processes in the bid and ask side remains uniformly bounded with respect to the state of the LOB. Assumption 3.3, which is the most important for the system's ergodicity, forces the individual queue sizes $|q_i|$ to decrease when they become larger than a certain threshold. From these assumptions, we obtain the following theorem proved in appendix for the Markov process $q(t)$.

Theorem 3.1. *When $u = d = 0$, under Assumptions 2.1, 3.1, 3.2, 3.3 and 3.4, the continuous-time Markov jump process $q(t)$ is non-explosive, V -uniformly ergodic and positive Harris recurrent.*

Consider now the embedded Markov chain $q(n)$ defined by $q(n) = q(J_n)$, where J_n is the time of the n -th jump, and $q(J_n)$ the state of the LOB after this event. The study of the embedded Markov chain is an important step in order to obtain the diffusivity of the price process in our setting. We write

$$a_i^*(q) = \frac{f_i^*(q)}{\sum_i [f_i^*(q) + g_i^*(q)]}, \quad b_i^*(q) = \frac{g_i^*(q)}{\sum_i [f_i^*(q) + g_i^*(q)]},$$

for the proportions of queue size increases and decreases, and replace Assumption 3.3 by the following one.

Assumption 3.5. *There exist $r > 0$ and $U > 1$ such that*

$$\begin{aligned} \overline{\lim}_{z \rightarrow 1^+} \sup_{(q,i): q_i > U, i \geq i_{bestask}} [a_i^*(q) - b_i^*(q) \frac{1 - G^{g,i,q}(z^{-1})}{G^{f,i,q}(z) - 1}] &< -r \\ \overline{\lim}_{z \rightarrow 1^+} \sup_{(q,i): q_i < -U, i \leq i_{bestbid}} [b_i^*(q) - a_i^*(q) \frac{1 - G^{f,i,q}(z^{-1})}{G^{g,i,q}(z) - 1}] &< -r. \end{aligned}$$

The following theorem is proved in appendix.

Theorem 3.2. *When $u = d = 0$, under Assumptions 2.1, 3.1, 3.2, 3.4 and 3.5, the embedded discrete-time Markov chain $q(n)$ is V -uniformly ergodic and positive Harris recurrent.*

⁷This assumption is not really mandatory but is realistic and technically quite convenient.

3.2 General case

We are now interested in the case where u and d are no longer fixed to 0. Recall that $q(n)$ represents the state of the LOB after the n -th event and $p_{ref}(n)$ is the reference price (the center of the LOB) after the n -th event. We thus consider here the process of reference price increments $c(n)$ (since the reference price itself is of course not ergodic), defined as the reference price change at the n -th event:

$$c(n) = p_{ref}(n) - p_{ref}(n-1),$$

and the embedded chain $Y(n) = (q(n), c(n))$, $n \in \mathbb{N}$, with $c(0)$ an artificial starting value. The process $Y(n)$ remains obviously Markovian. For some $z > 1$, let

$$V^z([q, c]) = \sum_{i=-K, i \neq 0}^K z^{|q_i| - U}.$$

We make two additional assumptions for the general case.

Assumption 3.6. *There exist $z > 1$ and $L^\pi > 0$ such that for Q^{inc} , Q^{dec} , Q_K , Q_{-K} four random variables such that $Q^{inc} \sim \pi^{inc}$, $Q^{dec} \sim \pi^{dec}$, $Q_K \sim \pi_K$ and $Q_{-K} \sim \pi_{-K}$:*

$$\mathbb{E}[V^z([Q^{inc}, c])] + \mathbb{E}[V^z([Q^{dec}, c])] + \mathbb{E}[z^{|Q_K| - U}] + \mathbb{E}[z^{|Q_{-K}| - U}] \leq L^\pi.$$

Assumption 3.7. *There exists a finite set $W \subset \Omega$ such that the upper bound of the proportion of reference price jumps in any order book state q is smaller than one on Ω/W :*

$$\sup_{q \in \Omega/W} \frac{u(q) + d(q)}{\sum_i [f_i^*(q) + g_i^*(q)] + u(q) + d(q)} < 1.$$

Assumption 3.6 is technical and imposes some regularities on the four distributions π^{inc} , π^{dec} , π_K and π_{-K} . Assumption 3.7 ensures that a reference price change is not the only possible event except for a finite number of LOB states. Under these assumptions, we have the following theorem proved in appendix on the ergodicity of the embedded chain $Y(n)$.

Theorem 3.3. *Under Assumptions 2.1, 3.1, 3.2, 3.4, 3.5, 3.6 and 3.7, the embedded discrete-time Markov chain $Y(n) = (q(n), c(n))$ is V -uniformly ergodic and positive Harris recurrent.*

4 Scaling limits

We are now interested in the scaling limit of the reference price process. Let J_i be the time of the i -th jump of the process. Let

$$N(t) = \inf\{n, J_n \leq t\}$$

be the number of events until time t , with the convention $\inf\{\emptyset\} = 0$. Let $Z(n)$ be the cumulative price change until the n -th event, that is $Z(0) = 0$ and for $n \geq 1$:

$$Z(n) = \sum_{i=1}^n c(i).$$

We have

$$Z(N(t)) = p_{ref}(t) - p_{ref}(0).$$

Thus it represents the reference price at time t recentered its starting value. We show in this section the diffusive behavior of $Z(N(t))$ as n tends to infinity.

Consider again the embedded chain $Y(n) = (q(n), c(n))$. From Theorem 3.3, $Y(n)$ is V-uniformly ergodic towards an invariant distribution π^* . We have the following theorem for the rescaled price process in event time $\hat{S}^{(n)}(t) := \frac{Z(\lfloor nt \rfloor)}{\sqrt{n}}$.

Theorem 4.1. *Under Assumptions 2.1, 3.1, 3.2, 3.4, 3.5, 3.6 and 3.7, if $\mathbb{E}_{\pi^*}[c(0)] = 0$, then the series*

$$\sigma^2 = \mathbb{E}_{\pi^*}[c_0^2] + 2 \sum_{n=1}^{\infty} \mathbb{E}_{\pi^*}[c_0 c_n], \quad (7)$$

converges absolutely. Furthermore, if $Y(0) \sim \pi^$, we have the following convergence in law in $D[0, \infty)$:*

$$\hat{S}^{(n)}(t) \xrightarrow{n \rightarrow \infty} \sigma B(t),$$

where $B(t)$ is a standard Brownian motion.

Proof. This theorem is a direct application of Theorem 19.1 in Billingsley (2009). Indeed, the sequence c_n is clearly stationary and ergodic in the sense of Billingsley (2009) (for example since it is stationary and mixing). Moreover, it has a finite second order moment and for all n , $\mathbb{E}_{\pi^*}[c_n] = \mathbb{E}_{\pi^*}[c_0] = 0$. \square

Theorem 4.1 shows that in event time, the large scale limit of the reference price process is a Brownian motion. However, the most relevant question is that of the large scale limit of the reference price in calendar time. Thus we now consider the process

$$\tilde{S}^{(n)}(t) = \frac{Z(N(nt))}{\sqrt{n}}.$$

To prove the diffusivity of $\tilde{S}^{(n)}(t)$, we need a last assumption which is a bound of the expected value on the waiting time between two events.

Assumption 4.1. *There exists some $m > 0$, such that*

$$\inf_{q \in \Omega} \left\{ \sum_i (f_i^*(q) + g_i^*(q)) + u(q) + d(q) \right\} > m.$$

Let τ_n be the inter-arrival time between the n -th and the $(n-1)$ -th jumps of the Markov process X . We then have the following theorem proved in appendix for the long term behavior of the reference price in calendar time.

Theorem 4.2. *Under Assumptions 2.1, 3.1, 3.2, 3.4, 3.5, 3.6, 3.7 and 4.1, the process $(q(n), c(n), \tau(n))$ is positive Harris recurrent. Furthermore, if $\mathbb{E}_{\pi^*}[c(0)] = 0$ and $Y(0) \sim \pi^*$, then*

$$\tilde{S}^{(n)}(t) \xrightarrow{n \rightarrow \infty} \frac{\sigma}{\sqrt{\mathbb{E}_{\pi^{**}}[\tau(1)]}} B(t),$$

*with π^{**} is the invariant distribution of $(q(n), c(n), \tau(n))$ and σ defined in (7).*

Theorem 4.2 discusses the scaling limit of the underlying reference price. However, the difference between this price and the more usual $p_{bestbid}(t)$, $p_{bestask}(t)$ or $p_{mid}(t)$ being bounded by $2K$, the same result applies replacing the reference price by any of those prices.

5 Some specific models

The Markovian setting proposed in this work allows us for a wide range of possibilities for modeling order book dynamics. The goal of this section is to give some natural and tractable examples of models, essentially already introduced in the literature, which can be seen as particular cases of our general framework. Together with the dynamics of the models, we provide sufficient conditions so that the assumptions made in the previous sections are satisfied in these specific models. Thus the ergodicity and diffusive scaling properties apply in all these models.

5.1 Best bid/best ask Poisson model (Cont and De Larrard (2013))

The basic idea of this first model, inspired by that introduced in Cont and De Larrard (2013), is to use a constant spread size (fixed to 1 tick) and to consider only two limits in the order book.

Example 3. *Poisson model with $K = 1$.*

- We take $K = 1$, $\theta^{reinit} = 1$ and assume that the functions f_i , g_i , u and d have the following forms, with $0 < \lambda < \mu < \infty$:

$$\begin{aligned} f_1(q, n) &= \lambda \mathbf{1}_{n=1} \\ g_1(q, n) &= \mu \mathbf{1}_{q_1 > 0} \mathbf{1}_{n=1} \\ f_{-1}(q, n) &= \mu \mathbf{1}_{q_{-1} < 0} \mathbf{1}_{n=1} \\ g_{-1}(q, n) &= \lambda \mathbf{1}_{n=1} \\ u(q) &= \theta \mathbf{1}_{q_1=0} \\ d(q) &= \theta \mathbf{1}_{q_{-1}=0}. \end{aligned}$$

- π^{inc} and π^{dec} satisfy Assumption 3.6 and

for any $q_{-1} > 0$, $q_1 \in \mathbb{Z}$,

$$\pi^{inc}(q_{-1}, q_1) = \pi^{dec}(q_{-1}, q_1) = 0,$$

for any $q_1 < 0$, $q_{-1} \in \mathbb{Z}$,

$$\pi^{inc}(q_{-1}, q_1) = \pi^{dec}(q_{-1}, q_1) = 0.$$

Note that here, the boundary distributions π_K and π_{-K} are no longer needed, since the order book reinitialization probability θ^{reinit} is set to one.

In this model, the role of p_{ref} is very close to that of p_{mid} , which splits the order book into two parts: the bid side (Q_{-1}) and the ask side (Q_1). The limit order insertion, cancellation and market order insertion processes are assumed to be independent Poisson processes. The size of these orders is assumed to be constant and p_{ref} jumps with rate 0 when none of the queues $Q_{\pm 1}$ is empty, with rate θ to the right side when Q_1 is empty, with rate θ to the left side when Q_{-1} is empty. When the value of θ is very large, the price jump is almost instantaneous as soon as one of the two queues becomes empty. In that case, this model becomes very close to that proposed in Cont and De Larrard (2013), where an infinite rate is used (note that the convergence of the rescaled price process can still be proved with some minor modifications in such case of infinite jump rate).

5.2 Poisson model with $K > 1$

It is natural to try to extend the previous Poisson model in order to include more queues in the order book and to allow for a spread size different from one tick. In such model, the role of p_{ref} is slightly different since it is not necessarily the mid price. Again, p_{ref} can be understood here as the underlying efficient price that determines the order arrival intensities at different price levels. Now buy/sell limit orders can be inserted both on the right side and on the left side of p_{ref} .

Example 4. *Poisson model with $K > 1$.*

- The functions f_i , g_i , u and d have the following forms, for $i = -K, \dots, K$:

$$\begin{aligned} f_i(q, 1) &= \lambda_i \mathbf{1}_{i > i_{bestbid}(q)} + \gamma_i \mathbf{1}_{i = i_{bestbid}(q)} \\ &\quad + \mu_i \mathbf{1}_{i \leq i_{bestbid}(q)} \mathbf{1}_{q_i < 0} \\ g_i(q, 1) &= \lambda_{-i} \mathbf{1}_{i < i_{bestask}(q)} + \gamma_{-i} \mathbf{1}_{i = i_{bestask}(q)} \\ &\quad + \mu_{-i} \mathbf{1}_{i \geq i_{bestask}(q)} \mathbf{1}_{q_i > 0} \\ f_i(q, n) &= 0, \text{ for } n > 1 \\ g_i(q, n) &= 0, \text{ for } n > 1 \\ u(q) &= \theta_{i_{bestask}(q)} \\ d(q) &= \theta_{-i_{bestbid}(q)}. \end{aligned}$$

- π^K , π^{-K} , π^{inc} and π^{dec} satisfy Assumption 3.6.
- For any $i, j \in \{-K, \dots, K\}$, $i < j$, we have

$$\begin{aligned} \mu_{-i} &> \lambda_i > 0 \\ 0 &\leq \theta_i \leq \theta_j. \end{aligned}$$

Limit orders, cancellations and market orders (which consume the quantities at the best offer limits) are modeled by independent Poisson processes with different intensities $\lambda_i^{buy/sell}$, $\mu_i^{buy/sell}$ and $\gamma_i^{buy/sell}$, depending on the distance from their target price to p_{ref} . We assume bid-ask symmetry in this model, that is $\lambda_i^{buy} = \lambda_{-i}^{sell}$, $\mu_i^{buy} = \mu_{-i}^{sell}$ and $\gamma_i^{buy} = \gamma_i^{sell}$, thus we omit the index buy/sell in the above equations. Remark that the intensity of the buy/sell market order flow at the best limit γ_i is a function of i , that is the position of the best limit with respect to the reference price. This allows us to model the fact that market participants have different behaviors towards the best limit, depending on their evaluation of the reference price.

The reference price jump dynamics is modeled by a sequence of increasing rates θ_i , $i \in \{-K, \dots, K\}$. This means that the larger the index of the best ask queue, the larger the probability of p_{ref} to increase and the smaller the index of the best bid queue, the larger the probability of p_{ref} to decrease. Note that in this model, we no longer assume any specific value for the reinitialization probability θ^{reinit} and use Assumption 3.6 to impose some properties on the boundary distributions π^K and π^{-K} and the initialization distributions π^{inc} and π^{dec} .

5.3 Zero-intelligence model

We now present a different way of extending the Poisson model with $K = 1$ in order to include more queues in the order book. This modeling approach where two reference prices are used is called Zero-intelligence model and is introduced in Smith, Farmer, Gillemot, and Krishnamurthy (2003). It is also considered in Abergel and Jedidi (2011) and is the basis of Cont, Stoikov, and Talreja (2010). We define $\phi(i, j)$ as the absolute distance (in number of ticks) between the queue Q_i and Q_j ($\phi(i, j) = |i - j|$) and make the three following assumptions.

Example 5. *Zero-intelligence model.*

- The functions f_i , g_i , u and d have the following forms, for $i = -K, \dots, K$:

$$\begin{aligned} f_i(q, 1) &= \lambda_{\phi(i, i_{bestbid}(q))} \mathbf{1}_{i > i_{bestbid}(q)} + \gamma \mathbf{1}_{i = i_{bestbid}(q)} + |q_i| \mu_{\phi(i, i_{bestask}(q))} \mathbf{1}_{i \leq i_{bestbid}(q)} \\ g_i(q, 1) &= \lambda_{\phi(i, i_{bestask}(q))} \mathbf{1}_{i < i_{bestask}(q)} + \gamma \mathbf{1}_{i = i_{bestask}(q)} + |q_i| \mu_{\phi(i, i_{bestbid}(q))} \mathbf{1}_{i \geq i_{bestask}(q)} \\ f_i(q, n) &= 0, \text{ for } n > 1 \\ g_i(q, n) &= 0, \text{ for } n > 1 \\ u(q) &= \theta_{i_{bestask}(q)} \\ d(q) &= \theta_{-i_{bestbid}(q)}. \end{aligned}$$

- π^K , π^{-K} , π^{inc} and π^{dec} satisfy Assumption 3.6.
- For any $i, j \in \{-K, \dots, K\}$, $\lambda_{\phi(i, j)} > 0$ and $\mu_{\phi(i, j)} > 0$.
- For any $i, j \in \{-K, \dots, K\}$, $i < j$, we have

$$0 \leq \theta_i \leq \theta_j.$$

In this model, p_{ref} is no longer an underlying efficient price determining the order arrival intensities. These intensities now depend on the positions of two different prices: $p_{bestbid}$ and $p_{bestask}$. Limit orders, cancellations and market orders are still described by independent Poisson processes. Buy/sell limit orders are inserted in the queues to the left/right side of the best ask/best bid price, with intensities depending on the distance between their price level and the best ask/best bid price ($\lambda_{\phi(i, i_{bestask/bestbid})}$); cancellations of buy/sell orders are sent to the queues on the left/right side of the best ask/best bid price, with intensities being linear functions of the queue sizes ($|q_i| \mu_{\phi(i, i_{bestask/bestbid})}$); market buy/sell orders are sent to the best ask/best bid queue, with intensity γ . The reference price p_{ref} now provides the center of the $2K$ dimensional moving frame representing the LOB's state and the same modeling approach as in Section 5.2 is used for its dynamics.

5.4 Queue-reactive model (Huang, Lehalle, and Rosenbaum (2015))

In Huang, Lehalle, and Rosenbaum (2015), the Queue-reactive model for order books is introduced. This model takes into account the influence of the order book's state in determining the order arrival intensities (in a much more general way than considering only the position of the best bid and best ask queues). The Queue-reactive model assumes that no buy/sell limit order can be inserted on the right/left side of p_{ref} and uses the following assumption instead of Equation (2) for the dynamics of the jumps of p_{ref} .

Assumption 5.1. *Whenever p_{mid} increases (resp. decreases), p_{ref} increases (resp. decreases) by α with probability θ , provided that $q_1 = 0$ (resp. $q_{-1} = 0$) at that moment. Therefore, changes in the value of p_{ref} are possibly triggered by one of the three following events:*

- *The insertion of a buy (resp. sell) limit order within the bid-ask spread while Q_1 (resp. Q_{-1} is empty).*
- *The cancellation of the last limit order at one of the best offer queues.*
- *A market order which consumes the last limit order at one of the best offer queues.*

With some minor modifications in the proof of Theorem 4.2, one can prove that the scaling limit of p_{ref} in the Queue-reactive model is a Brownian motion. As explained above, in this model, changes of p_{ref} are triggered by events that modify the mid price. Here we propose a slightly modified version of the Queue-reactive model⁸ by considering the following four assumptions (note that the state space Ω is reduced in that case to $\Omega^* := \{q \in \Omega, q_i \mathbf{1}_{i < 0} \leq 0, q_i \mathbf{1}_{i > 0} \geq 0\}$).

Example 6. *Queue-reactive type model.*

- *The functions f_i, g_i, u and d have the following forms, for $i = -K, \dots, K$:*

$$\begin{aligned} f_i(q, 1) &= \lambda_{|i|}(q_i) \mathbf{1}_{i > 0} + \mu_{|i|}(-q_i) \mathbf{1}_{i < 0} \\ g_i(q, 1) &= \lambda_{|i|}(-q_i) \mathbf{1}_{i < 0} + \mu_{|i|}(q_i) \mathbf{1}_{i > 0} \\ f_i(q, n) &= 0, \text{ for } n > 1 \\ g_i(q, n) &= 0, \text{ for } n > 1 \\ u(q) &= \theta \mathbf{1}_{q_1=0} \\ d(q) &= \theta \mathbf{1}_{q_{-1}=0}, \end{aligned}$$

with $\lambda_{|i|}$ and $\mu_{|i|}$ non-negative functions defined on \mathbb{N} , with $\mu_{|i|}(0) = 0$.

- *We have*

$$\sup_{i \in \{1, \dots, K\}, q_i \in \mathbb{N}} (\lambda_i(q_i)) < L < \infty.$$

- *There exist $r > 0$ and $U > 1$ such that for any $q_i > U$ and any $i \in \{1, \dots, K\}$:*

$$\lambda_i(q_i) - \mu_i(q_i) < -r.$$

- *There exists $m' > 0$ such that for any $i \in \{1, \dots, N\}$:*

$$\inf_{q_i \in \mathbb{N}} [\lambda_i(q_i) + \mu_i(q_i)] > m'.$$

Compared with Assumption 5.1, changes in p_{ref} are now driven by the relative position of the mid price in the current order book state. Nevertheless, we can see that the two approaches are quite similar. In this model, p_{ref} always stays between $p_{bestask}$ and $p_{bestbid}$ (since $\mu_{|i|}(0) = 0$ implies that the queue sizes on the left/right side of p_{ref} never become positive/negative). Such

⁸Model I in Huang, Lehalle, and Rosenbaum (2015) is used to describe the queue dynamics during constant reference price periods.

model gives us a much larger choice on the intensity functions $\lambda_{|i|}$ and $\mu_{|i|}$ than when assuming Poisson flows. Furthermore, with enough data points, these functions can be estimated in a non-parametric way, as done in Huang, Lehalle, and Rosenbaum (2015). Finally the state-dependent approach provides us very interesting insights about the way the order book state influences market participants decisions and the mechanism making the empirical distribution of the order book arise from these decisions, see Huang, Lehalle, and Rosenbaum (2015).

6 Conclusion

In this work, we extend the order book modeling approach proposed in Huang, Lehalle, and Rosenbaum (2015) to a more general Markovian framework, allowing to take into account most relevant features of LOB dynamics such as:

- Dependencies between the order arrival processes and the LOB state.
- Endogenous movements of the underlying efficient price and influence of the LOB state on its dynamics.
- Exogenous movements of the underlying efficient price.
- Randomness in the size of the orders.

The ergodicity of the LOB system and the diffusive limit of the rescaled price process are established under general assumptions. Finally, to illustrate the usefulness and the relevance of our approach, several examples of classical models which can be seen as particular cases of our general framework are presented.

To get a fully satisfying model, a last step would probably be to allow for a non-Markovian component in the market order flow (since the Markov assumption is probably quite reasonable for the limit order and cancellation flows). This can for example be done using self-exciting processes, as in Abergel and Jedidi (2015). However, except for very specific cases (exponential Hawkes processes for example), adding such non-Markovian component would certainly require revising completely the mathematical approach to the model.

Acknowledgements

We thank Frédéric Abergel for helpful discussions.

7 Appendix

7.1 Proof of Theorem 3.1

Let us denote by $\tilde{\mathcal{Q}}$ the infinitesimal generator matrix of $q(t)$ when $u = d = 0$. The infinite matrix $\tilde{\mathcal{Q}}$ has the following form:

$$\begin{aligned}\tilde{\mathcal{Q}}_{q, q+ne_i} &= f_i(q, n) \\ \tilde{\mathcal{Q}}_{q, q-ne_i} &= g_i(q, n) \\ \tilde{\mathcal{Q}}_{q, q} &= - \sum_{q \in \Omega, q' \neq q} \tilde{\mathcal{Q}}_{q, q'} \\ \tilde{\mathcal{Q}}_{q, q'} &= 0, \text{ otherwise.}\end{aligned}$$

For some $1 < z \leq z^*$ (recall that z^* is defined in Assumption 3.2), let us consider the function

$$V(q) = \sum_{i=-K, i \neq 0}^K z^{|q_i| - U}.$$

Since $q_i \geq 0$ for $i \geq i_{bestask}$, $q_i \leq 0$ for $i \leq i_{bestbid}$ and $q_i = 0$ for $i \in (i_{bestbid}, i_{bestask})$, we have

$$\begin{aligned}\tilde{\mathcal{Q}}V(q) &:= \sum_{q' \neq q} \tilde{\mathcal{Q}}_{qq'} [V(q') - V(q)] \\ &= \sum_{i \leq i_{bestbid}} \sum_{n=1}^{\infty} [(z^{-q_i+n-U} - z^{-q_i-U})g_i(q, n) + (z^{|q_i+n|-U} - z^{-q_i-U})f_i(q, n)] \\ &\quad + \sum_{i \geq i_{bestask}} \sum_{n=1}^{\infty} [(z^{q_i+n-U} - z^{q_i-U})f_i(q, n) + (z^{|q_i-n|-U} - z^{q_i-U})g_i(q, n)] \\ &\quad + \sum_{i \in (i_{bestbid}, i_{bestask})} \sum_{n=1}^{\infty} [(z^{n-U} - z^{-U})f_i(q, n) + (z^{n-U} - z^{-U})g_i(q, n)].\end{aligned}$$

Then by Assumption 3.1,

$$\begin{aligned}\tilde{\mathcal{Q}}V(q) &= \sum_{i \leq i_{bestbid}} \sum_{n=1}^{\infty} [(z^{-q_i+n-U} - z^{-q_i-U})g_i(q, n) + (z^{-q_i-n-U} - z^{-q_i-U})f_i(q, n)] \\ &\quad + \sum_{i \geq i_{bestask}} \sum_{n=1}^{\infty} [(z^{q_i+n-U} - z^{q_i-U})f_i(q, n) + (z^{q_i-n-U} - z^{q_i-U})g_i(q, n)] \\ &\quad + \sum_{i \in (i_{bestbid}, i_{bestask})} \sum_{n=1}^{\infty} [(z^{n-U} - z^{-U})f_i(q, n) + (z^{n-U} - z^{-U})g_i(q, n)] \\ &= \sum_{i \leq i_{bestbid}} z^{-q_i-U} \sum_{n=1}^{\infty} (z^n - 1) [g_i(q, n) - \frac{f_i(q, n)}{z^n}] \\ &\quad + \sum_{i \geq i_{bestask}} z^{q_i-U} \sum_{n=1}^{\infty} (z^n - 1) [f_i(q, n) - \frac{g_i(q, n)}{z^n}] \\ &\quad + \sum_{i \in (i_{bestbid}, i_{bestask})} z^{-U} \sum_{n=1}^{\infty} (z^n - 1) [f_i(q, n) + g_i(q, n)].\end{aligned}$$

Using the definition of $G^{f,i,q}(z)$, $G^{g,i,q}(z)$, $f_i^*(q)$ and $g_i^*(q)$, we get

$$\begin{aligned}\tilde{Q}V(q) &= \sum_{i \leq i_{bestbid}} z^{-q_i-U} [g_i^*(q)(G^{g,i,q}(z) - 1) - f_i^*(q)(1 - G^{f,i,q}(z^{-1}))] \\ &+ \sum_{i \geq i_{bestask}} z^{q_i-U} [f_i^*(q)(G^{f,i,q}(z) - 1) - g_i^*(q)(1 - G^{g,i,q}(z^{-1}))] \\ &+ \sum_{i \in (i_{bestbid}, i_{bestask})} z^{-U} [f_i^*(q)(G^{f,i,q}(z) - 1) + g_i^*(q)(G^{g,i,q}(z) - 1)].\end{aligned}$$

Moreover, since for $z > 1$ we have $G^{f/g,i,q}(z^{-1}) < 1$ and $G^{f/g,i,q}(z) > 1$, we obtain

$$\begin{aligned}\tilde{Q}V(q) &\leq \sum_{i: i \leq i_{bestbid}, q_i < -U} z^{-q_i-U} (G^{g,i,q}(z) - 1)(g_i^*(q) - f_i^*(q)) \frac{1 - G^{f,i,q}(z^{-1})}{G^{g,i,q}(z) - 1} \\ &+ \sum_{i: i \geq i_{bestask}, q_i > U} z^{q_i-U} (G^{f,i,q}(z) - 1)(f_i^*(q) - g_i^*(q)) \frac{1 - G^{g,i,q}(z^{-1})}{G^{f,i,q}(z) - 1} \\ &+ \sum_{i: i \leq i_{bestbid}, q_i \geq -U} g_i^*(q) G^{g,i,q}(z) + \sum_{i: i \geq i_{bestask}, q_i \leq U} f_i^*(q) G^{f,i,q}(z) \\ &+ \sum_{i \in (i_{bestbid}, i_{bestask})} [f_i^*(q) G^{f,i,q}(z) + g_i^*(q) G^{g,i,q}(z)] \\ &\leq \sum_{i: i \leq i_{bestbid}, q_i < -U} z^{-q_i-U} (G^{g,i,q}(z) - 1)(g_i^*(q) - f_i^*(q)) \frac{1 - G^{f,i,q}(z^{-1})}{G^{g,i,q}(z) - 1} \\ &+ \sum_{i: i \geq i_{bestask}, q_i > U} z^{q_i-U} (G^{f,i,q}(z) - 1)(f_i^*(q) - g_i^*(q)) \frac{1 - G^{g,i,q}(z^{-1})}{G^{f,i,q}(z) - 1} \\ &+ \sum_i [f_i^*(q) G^{f,i,q}(z) \mathbf{1}_{i > i_{bestbid}} + g_i^*(q) G^{g,i,q}(z) \mathbf{1}_{i < i_{bestask}}].\end{aligned}$$

Now note that

$$\sup_q [f_i^*(q) G^{f,i,q}(z) \mathbf{1}_{i > i_{bestbid}} + g_i^*(q) G^{g,i,q}(z) \mathbf{1}_{i < i_{bestask}}]$$

is an increasing function of z . Thus by Assumption 3.2, we can find $z' > 1$ such that for any $z \leq z'$, $q \in \Omega$ and $i \in [-K, \dots, K]$:

$$(f_i^*(q) G^{f,i,q}(z) \mathbf{1}_{i > i_{bestbid}} + g_i^*(q) G^{g,i,q}(z) \mathbf{1}_{i < i_{bestask}}) < L.$$

From Assumption 3.3, we see that we can find some \tilde{z} with $1 < \tilde{z} \leq z'$ such that for any (q, i) , $q_i < -U$, $i \leq i_{bestbid}$, for any $1 < z \leq \tilde{z}$:

$$g_i^*(q) - f_i^*(q) \frac{1 - G^{f,i,q}(z^{-1})}{G^{g,i,q}(z) - 1} < -r, \quad (8)$$

and for any (q, i) , $q_i > U$, $i \geq i_{bestask}$, for any $1 < z \leq \tilde{z}$:

$$f_i^*(q) - g_i^*(q) \frac{1 - G^{g,i,q}(z^{-1})}{G^{f,i,q}(z) - 1} < -r. \quad (9)$$

Thus taking z in the definition of the function V satisfying $2Kz^{-U} > 1$ (so that the function V is coercive) and $1 < z \leq \tilde{z}$, we obtain

$$\begin{aligned}
\tilde{Q}V(q) &\leq -r \sum_{i: i \leq i_{bestbid}, q_i < -U} z^{-q_i - U} (G^{g,i,q}(z) - 1) \\
&\quad -r \sum_{i: i \geq i_{bestask}, q_i > U} z^{q_i - U} (G^{f,i,q}(z) - 1) + 2KL. \\
&\leq -rB_g(z) \sum_{i: i \leq i_{bestbid}, q_i < -U} z^{-q_i - U} \\
&\quad -rB_f(z) \sum_{i: i \geq i_{bestask}, q_i > U} z^{q_i - U} + 2KL.
\end{aligned}$$

By Assumption 3.4, $B := \min(B_g(z), B_f(z)) > 0$. Therefore we get

$$\begin{aligned}
\tilde{Q}V(q) &\leq -rB \sum_{i: |q_i| > U} z^{|q_i| - U} + 2KL \\
&\leq -rB \sum_i z^{|q_i| - U} + 2K(L + rB) \\
&= -rBV(q) + 2K(L + rB).
\end{aligned}$$

Finally, remark that in our setting, any compact set included in Ω is finite. A singleton being a petite set and a finite union of singletons remaining a petite set, see Proposition 5.5.5 in Meyn and Tweedie (2009), we get that all the compact sets are petite. Therefore by Theorem 6.1 in Meyn and Tweedie (1993), $q(t)$ is non-explosive and V-uniformly ergodic. Furthermore, by Theorem 4.2 in Meyn and Tweedie (1993) it is positive Harris recurrent.

7.2 Proof of Theorem 3.2

For some $1 < z \leq z^*$, set again

$$V(q) = \sum_{i=-K, i \neq 0}^K z^{|q_i| - U}.$$

We write $\tilde{\mathcal{P}}_{q,q'}$ the transition probability from q to q' . In the same way as in the preceding proof, we have

$$\begin{aligned}
\Delta V(q) &:= \sum_{q' \in \Omega} \tilde{\mathcal{P}}_{q,q'} (V(q') - V(q)) \\
&= \sum_{i \leq i_{bestbid}} z^{-q_i - U} [b_i^*(q)(G^{g,i,q}(z) - 1) - a_i^*(q)(1 - G^{f,i,q}(z^{-1}))] \\
&\quad + \sum_{i \geq i_{bestask}} z^{q_i - U} [a_i^*(q)(G^{f,i,q}(z) - 1) - b_i^*(q)(1 - G^{g,i,q}(z^{-1}))] \\
&\quad + \sum_{i \in (i_{bestbid}, i_{bestask})} z^{-U} [a_i^*(q)(G^{f,i,q}(z) - 1) + b_i^*(q)(G^{g,i,q}(z) - 1)].
\end{aligned}$$

Following the same method as in the proof of Theorem 3.1, we can easily find $1 < z' \leq z^*$ and $B > 0$ such that taking $z = z'$ in the definition of V , it is coercive and we get

$$\Delta V(q) \leq -rV(q) + B.$$

Now define the set $C := \{q, rV(q) \leq 2B\}$, C is obviously a finite set and is therefore petite. Furthermore, we have

$$\Delta V(q) \leq -\frac{r}{2}V(q) + B\mathbf{1}_{q \in C}.$$

Thus by Theorem 16.1.2 in Meyn and Tweedie (2009), $q(n)$ is V -uniformly ergodic.

Eventually, the fact that the chain is positive Harris recurrent is deduced from Theorem 9.1.8 together with Theorem 15.0.1 in Meyn and Tweedie (2009).

7.3 Proof of Theorem 3.3

For ease of notation, we write V instead of V^z . Let Q^{inc} , Q^{dec} , Q_K , Q_{-K} be four random variables such that $Q^{inc} \sim \pi^{inc}$, $Q^{dec} \sim \pi^{dec}$, $Q_K \sim \pi_K$ and $Q_{-K} \sim \pi_{-K}$. We define

$$\begin{aligned} u^*(q) &= \frac{u(q)}{\sum_i [f_i^*(q) + g_i^*(q)] + u(q) + d(q)} \\ d^*(q) &= \frac{d(q)}{\sum_i [f_i^*(q) + g_i^*(q)] + u(q) + d(q)} \\ n^*(q) &= \frac{\sum_i [f_i^*(q) + g_i^*(q)]}{\sum_i [f_i^*(q) + g_i^*(q)] + u(q) + d(q)} \\ \mathbb{E}_K &= \mathbb{E}[z^{Q_K - U}] \\ \mathbb{E}_{-K} &= \mathbb{E}[z^{Q_{-K} - U}] \\ \mathbb{E}_{\pi^{inc}} &= \mathbb{E}[V([Q^{inc}, c])] \\ \mathbb{E}_{\pi^{dec}} &= \mathbb{E}[V([Q^{dec}, c])]. \end{aligned}$$

Remarking that $V([q, c])$ does not depend on c , we write from now on $V(q)$ instead of $V([q, c])$. Moreover, we set $\mathcal{P}_{[q, c], [q', c']}$ as the transition probability from state $[q, c]$ to state $[q', c']$ and $\tilde{\mathcal{P}}_{q, q'}$ as the transition matrix of the embedded chain $q(n)$ when $u = d = 0$. Using the form of the infinitesimal generator \mathcal{Q} , we deduce

$$\begin{aligned} \Delta V([q, c]) &:= \sum_{(q', c') \in \Omega \times \{-\alpha, \alpha\}} \mathcal{P}_{[q, c], [q', c']} (V(q') - V(q)) \\ &= n^*(q) \sum_{q' \in \Omega} \tilde{\mathcal{P}}_{q, q'} (V(q') - V(q)) \\ &\quad + u^*(q) [(1 - \theta^{reinit})(\mathbb{E}_K - z^{|q_{-K}| - U}) + \theta^{reinit}(\mathbb{E}_{\pi^{inc}} - V(q))] \\ &\quad + d^*(q) [(1 - \theta^{reinit})(\mathbb{E}_{-K} - z^{|q_K| - U}) + \theta^{reinit}(\mathbb{E}_{\pi^{dec}} - V(q))]. \end{aligned}$$

By Assumption 3.6, we have

$$\Delta V([q, c]) \leq n^*(q) \sum_{q' \in \Omega} \tilde{\mathcal{P}}_{q, q'} (V(q') - V(q)) + 2L^K + 2L^\pi.$$

Then as in the proof of Theorem 3.2, we can find $1 < z' \leq z^*$ and $B > 0$ such that taking $z = z'$ in the definition of V , it is coercive and we get

$$\Delta V(q) \leq -n^*(q)rV(q) + Bn^*(q) + 2L^K + 2L^\pi.$$

Moreover, Assumption 3.7 ensures that for any q except those belonging to the finite set W , $n^*(q) > M$ with $M \in (0, 1]$. Consequently,

$$\Delta V(q) \leq -MrV(q) + B + 2L^K + 2L^\pi + \nu \mathbf{1}_{[q,c] \in W \times \{-\alpha, \alpha\}},$$

with $\nu > 0$. Now define the set $C := \{[q, c], MrV(q) \leq 2B + 4L^K + 4L^\pi\}$. Being finite, $C \cup W$ is a petite set and we have

$$\Delta V([q, c]) \leq -\frac{Mr}{2}V([q, c]) + (B + 2L^K + 2L^\pi + \nu)\mathbf{1}_{[q,c] \in C \cup W}.$$

Hence from Theorem 16.1.2 in Meyn and Tweedie (2009), $Y(n)$ is V-uniformly ergodic.

Eventually, the fact that the chain is positive Harris recurrent is deduced from Theorem 9.1.8 together with Theorem 15.0.1 in Meyn and Tweedie (2009).

7.4 Proof of Theorem 4.2

7.4.1 Preliminary lemma

We start with the following preliminary lemma.

Lemma 7.1. *For the Markov chain $(q(n), c(n), \tau(n))$, the Cartesian product of any finite set included in $\Omega \times \{-\alpha, \alpha\}$ and \mathbb{R}^+ is petite.*

Proof. We first show that for any $q \in \Omega$ and $c \in \{-\alpha, \alpha\}$, the set $q \times c \times \mathbb{R}^+$ is petite (actually small). We define the measure $\nu_{q,c,\tau}$ which is so that for any $B \in \mathcal{B}(\Omega \times \{-\alpha, \alpha\} \times \mathbb{R}^+)$, $\nu_{q,c,\tau}(B)$ is the transition probability from $[q, c, \tau]$ to B in a single step:

$$\nu_{q,c,\tau}(B) = \mathcal{P}_{[q,c,\tau],B}.$$

Recall that in our framework, the transition probabilities from $(q(n), c(n), \tau(n))$ depend only on the value of $q(n)$. So we can write $\nu_{q,c,\tau}(B)$ as $\nu_q(B)$. In the sense of Equation (5.43) in Meyn and Tweedie (2009), the transition probability $\mathcal{P}_{[q,c,\tau],B}$ can be seen as a sampling kernel for the Markov chain $(q(n), c(n), \tau(n))$, using the Dirac measure at point 1 on \mathbb{Z}^+ as sampling measure. Moreover, for any $\tau \in \mathbb{R}^+$ and any $B \in \mathcal{B}(\Omega \times \{-\alpha, \alpha\} \times \mathbb{R}^+)$, we have

$$\mathcal{P}_{[q,c,\tau],B} \geq \nu_q(B).$$

Therefore the set $q \times c \times \mathbb{R}^+$ is petite. Then, as the union of two petite sets remains petite, see Proposition 5.5.5 in Meyn and Tweedie (2009), we have the result. \square

7.4.2 A law of large numbers

In the next proposition, we give a law of large numbers for the inter-arrival times, which is the key element to establish the diffusive behavior of the price in calendar time. Within the proof of this proposition, we show that $(q(n), c(n), \tau(n))$ is positive Harris recurrent.

Proposition 7.1. *Let τ_i be the inter-arrival time between the i -th and the $i-1$ -th jumps of the Markov process X . Under Assumptions 3.1, 3.2, 3.4, 3.5, 3.6, 3.7 and 4.1, almost surely, we have*

$$\frac{1}{n} \sum_{i=1}^n \tau_i \xrightarrow[n \rightarrow \infty]{} \mathbb{E}_{\pi^{**}}[\tau(1)],$$

with π^{**} the invariant distribution of $(q(n), c(n), \tau(n))$.

Proof. First we show that the Markov chain $(q(n), c(n), \tau(n))$ is positive Harris recurrent. For the Markov chain $(q(n), c(n))$, we have already proved that a coercive function V can be found, such that the following drift condition is satisfied for some $a > 0$ and $L < \infty$:

$$\Delta V([q, c]) \leq -aV([q, c]) + L.$$

Now, for the Markov chain $(q(n), c(n), \tau(n))$, take

$$V^*([q, c, t]) = V([q, c]) + t.$$

With obvious notation, we have

$$\begin{aligned} \Delta V^*([q, c, t]) &= \Delta V([q, c]) + \mathbb{E}_{[q, c, t]}[\tau] - t \\ &\leq -aV([q, c]) + L + \mathbb{E}_x[\tau] - t. \end{aligned}$$

Taking $a' = \min(a, 1)$ and $L' = L + 1/m$, using Assumption 4.1, we get

$$\begin{aligned} \Delta V^*([q, c, t]) &\leq -aV([q, c]) + L + 1/m - t \\ &\leq -a'[V([q, c]) + t] + L + 1/m \\ &\leq -a'V^*([q, c, t]) + L'. \end{aligned}$$

Now let $C = \{([q, c, t]), a'V^*([q, c, t]) \leq 2L'\}$. We have

$$\Delta V^*([q, c, t]) \leq -\frac{a'}{2}V^*([q, c, t]) + L'\mathbf{1}_{[q, c, t] \in C}.$$

According to Lemma 7.1, the set C is petite. Thus we can apply Theorem 15.0.1 in Meyn and Tweedie (2009) to deduce that the Markov chain $(q(n), r(n), \tau(n))$ is positive recurrent and thus admits an invariant measure.

Now remark that the function V^* is unbounded off petite sets (using Lemma 7.1 together with the fact that any subset of a petite set is itself petite). Consequently, Theorem 9.1.8 in Meyn and Tweedie (2009) enables us to obtain that the Markov chain $(q(n), c(n), \tau(n))$ is Harris recurrent. Therefore it is positive Harris recurrent.

We end the proof thanks to Theorem 17.0.1 from Meyn and Tweedie (2009). \square

7.4.3 End of the proof of Theorem 4.2

We have

$$\tilde{S}^{(n)}(t) = \frac{Z(\lfloor tn/\mathbb{E}_{\pi^{**}}[\tau(1)] \rfloor)}{\sqrt{n}} + \left(\frac{Z(N(nt))}{\sqrt{n}} - \frac{Z(\lfloor tn/\mathbb{E}_{\pi^{**}}[\tau(1)] \rfloor)}{\sqrt{n}} \right).$$

According to Proposition 7.1, the sequence of processes $N(nt)/n$ converges to $t/\mathbb{E}_{\pi^{**}}[\tau(1)]$. Moreover, the limit of $Z(\lfloor tn/\mathbb{E}_{\pi^{**}}[\tau(1)] \rfloor)/\sqrt{n}$ is continuous. Thus, using Skorohod representation theorem together with continuity properties in Skorohod topology, see Proposition VI.2.1 in Jacod and Shiryaev (2003), we get that the second term on the right hand side of the above equality tends to zero. Finally, from Theorem 4.1, we get

$$\frac{Z(\lfloor tn/\mathbb{E}_{\pi^{**}}[\tau(1)] \rfloor)}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} \frac{\sigma}{\sqrt{\mathbb{E}_{\pi^{**}}[\tau(1)]}} B_t.$$

Combining these two results, we obtain the weak convergence of the rescaled price process to a Brownian motion with variance $\sigma^2/\mathbb{E}_{\pi^{**}}[\tau(1)]$, which concludes the proof.

References

- Abergel, F. and Jedidi, A. (2011), “A mathematical approach to order book modelling,” in *Econophysics of Order-driven Markets*, Springer, pp. 93–107.
- (2015), “Long time behaviour of a Hawkes process-based limit order book,” *Preprint*.
- Bayer, C., Horst, U., and Qiu, J. (2015), “A functional limit theorem for limit order books with state dependent price dynamics,” *Preprint arXiv:1405.5230*.
- Billingsley, P. (2009), *Convergence of probability measures*, vol. 493, John Wiley & Sons.
- Bouchaud, J.-P., Mézard, M., and Potters, M. (2002), “Statistical properties of stock order books: empirical results and models,” *Quantitative finance*, 2, 251–256.
- Cont, R. and De Larrard, A. (2013), “Price dynamics in a Markovian limit order market,” *SIAM Journal on Financial Mathematics*, 4, 1–25.
- Cont, R., Stoikov, S., and Talreja, R. (2010), “A stochastic model for order book dynamics,” *Operations research*, 58, 549–563.
- Foucault, T. (1999), “Order flow composition and trading costs in a dynamic limit order market,” *Journal of Financial markets*, 2, 99–134.
- Huang, W., Lehalle, C.-A., and Rosenbaum, M. (2015), “Simulating and analyzing order book data: The queue-reactive model,” *Journal on the American Statistical Association*, 110, 107–122.
- Jacod, J. and Shiryaev, A. N. (2003), *Limit theorems for stochastic processes*, Springer-Verlag, Berlin.
- Lachapelle, A., Lasry, J.-M., Lehalle, C.-A., and Lions, P.-L. (2014), “Efficiency of the Price Formation Process in Presence of High Frequency Participants: a Mean Field Game analysis,” *Preprint arXiv:1305.6323*.
- Lakner, P., Reed, J., and Stoikov, S. (2013), “High Frequency Asymptotics for the Limit Order Book,” *Preprint*.
- Meyn, S. P. and Tweedie, R. L. (1993), “Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes,” *Advances in Probability*, 25, 518–548.
- (2009), *Markov chains and stochastic stability*, Cambridge University Press.
- Parlour, C. A. (1998), “Price dynamics in limit order markets,” *Review of Financial Studies*, 11, 789–816.
- Roşu, I. (2009), “A dynamic model of the limit order book,” *Review of Financial Studies*, 22, 4601–4641.
- Smith, E., Farmer, D. J., Gillemot, L., and Krishnamurthy, S. (2003), “Statistical theory of the continuous double auction,” *Quantitative Finance*, 3, 481–514.